Examination of Metalinguistic Skills in 3- to 8-Year Old Children with Low Language Abilities

A THESIS SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA BY

Amanda Kriese

IN PARTIAL FULFILLMENT OF THE REQUIERMENTS FOR THE DEGREE OF MASTER OF ARTS

Dr. Lizbeth Finestack

May 2019

Abstract

The current study examined the metalinguistic skills of children aged 3- to 8-years of age with low language abilities. The study aimed to determine if children with low language demonstrate weaknesses on metalinguistic tasks compared to typically developing peers. The study participants completed a battery of metalinguistic tasks to examine the their language knowledge and ability to process semantic, syntactic, and morphologic features of language. Overall, *t*-tests revealed that children who present with low language also demonstrate decreased metalinguistic skills when compared to peers of the same age. Additionally, results indicated that performance increased with age, and varied between tasks. Results aid in understanding language development and can be used to guide future interventions.

Table of Contents

List of Tables

List of Figures

Metalinguistic skills reflect awareness of grammatical, semantic, and/or morphological properties that make up language, as well as the ability to create and manipulate utterances using appropriate linguistic properties. Fujiki et.al (1986) and Bialystok (2001) defined metalinguistic skills as having two main components: knowledge of linguistic structures and the ability to process linguistic features. Knowledge of linguistic structures means that the child is aware of grammatical structures, that words mean something, and that morphemes can alter meaning. The ability to process linguistic features allows for the child to represent and attend to arbitrary ideas as meaningful utterances to support language use and development (Bialystok, 2001). Given the link between language and metalinguistic skills, it is likely that children with language learning weaknesses also have compromised metalinguistic skills. Children with language disorders are characterized by decreased skills related to syntax, morphology, semantics, word finding, pragmatics, discourse, as well as verbal learning, which may manifest from biological, environmental, or developmental causes (Bishop et. al, 2017). Functionally, children with language disorders are at risk for difficulties with language across their lifetime, leading to poor academic outcomes and general challenges with communication which may negatively impact relationships (Bishop, 2017).

While it is suggested that children with language impairment demonstrate decreased metalinguistic awareness, the majority of research has focused on the

metalinguistic skills of children who are bilingual (Kamhi et al., 1985; Fujiki et al., 1987; Redmond & Rice, 2001). Moreover, of the studies that have examined the metalinguistic skills of children with language learning weaknesses, most have focused on a single metalinguistic skill, limiting the ability to determine if the link is consistent across all metalinguistic skills or if it is limited to specific linguistic properties. Thus, the current study aims to expand on previous findings to understand the relationship between metalinguistic awareness and language abilities.

Metalinguistic Skills

In conjunction with language skills and knowledge, Bialystok (1987) asserts that to demonstrate metalinguistic skills, children must develop particular cognitive skills, including the ability to utilize control and problem-solving to analyze problems. Further stating that metalinguistic tasks are highly complex because they require a combination of cognitive and language demands. Cairns et. al (2006) adds that an individual must be able to reflect on language structures and components. Thus, children who experience difficulty with knowledge of language structures and processing of linguistic features are likely to have relatively weak metalinguistic skills (Fujiki et al., 1987).

Bialystok (1987) examined the metalinguistic skills of children who are bilingual. Bialystok noted that children who are bilingual are likely to develop skills needed to process linguistic information at a higher level compared to monolingual children; however, children who are bilingual do not have significantly higher knowledge of linguistic features. Thus, higher levels of processing ability and control are required to

use a second language. Bialystok's work (2001) also suggests that metalinguistic skills are task dependent and develop continuously as linguistic skills and demands develop. Due to task specificity, Bialystok (2001) argues that children who are bilingual do not have overall higher language abilities than children who are monolingual, but rather they have greater abilities to process and manipulate linguistic features, skills in which are important for children to successfully communicate across two languages. For children who are bilingual, sustaining interactions in a second language requires that the child maintain high levels of processing across varying conversational situations, as they continue to assess which language structures to use throughout the conversation.

Bialystok (1987) further suggests a link between linguistic abilities and performance on metalinguistic tasks, stating that language is made of multiple components, including spoken language, literacy, and metalinguistic skills. These components are also related, as metalinguistic skills are the ability to take a mental representation and form sentences that align with syntactic, semantic, and morphological rules for spoken language. This concept is supported by children who are bilingual outperforming children who are monolingual on their ability to process linguistic features for spoken language.

Bialystok (1987) reports on two metalinguistic tasks that have identified differences between metalinguistic skills in children who are bilingual to children who are monolingual, including a grammatical judgement task and a word swap task. The grammatical judgement task assesses the child's ability to separate meaning from

sentence structure, requiring the ability to process the linguistic features. For this task, an examiner provides children sentences that are grammatically correct but with incorrect meaning (e.g., Apples grow on noses.), grammatically incorrect with correct meaning (e.g., The girl are nice.), grammatically incorrect with incorrect meaning (e.g., The boot are look.), or grammatically correct with correct meaning (e.g., She is a teacher.). The examiner asks the child if the sentence is correct only based on the grammar, not the meaning. The word swap task also demands that the child process linguistic features separate from meaning. For this task, the examiner tells the child that they are playing a game and in which the names of the sun and moon are going to be swapped, but not the meaning. The examiner then shows the child a picture of the sun and asked what it is called (Piaget, 1929). If the child understands the task, the child should respond, "moon." When asked, "what would the sky look like when you see this?", the child should then respond, "blue."

Bialystok (1987) recognizes that there have been a variety of metalinguistic tasks used in previous studies and that each task provides unique information regarding metalinguistic skills. Other tasks used by researchers include a word manipulation task and a morpheme production task. In the word manipulation task (Chaney, 1992), the examiner explains that they are making up a new language and that that they will use new, novel names to represent common objects. For example, "carrot" will be called "gok." The examiner then asks the child questions to assess the child's understanding that a word represents a specific meaning, but the meaning and word are not dependent on

one another. For example, the examiner may ask questions such as "Are goks orange?" or "Do goks talk?". The morpheme production task assesses the child's ability to apply known morphological rules to novel words. An example of such a task is the Wug task created by Berko-Gleason (1958). During this task, the examiner shows the child a picture of a novel creature or object and labels the picture by saying, "This is a wug.". The examiner then shows the child a picture with two of the novel creatures and says, "Now there is another one, there are two _____." The target would be for the child to add the plural morpheme and respond, "wugs."

Metalinguistic Skills Children with Language Disorders

Studies examining the metalinguistic skills of children with language impairments have found decreased metalinguistic skills compared to children with typical development. Findings were consistent across tasks that assess syntax and morphological structure using segmenting, grammatical judgement, and word description tasks (Kamhi et. al, 1985; Fujiki et. al, 1987; Redmond & Rice, 2001). Kamhi et. al (1985) examined the metalinguistic skills of children aged 3 to 6 years of age with language impairment. Participants included a total of 45 children. Of these children, 15 were identified as having a language impairment by a school speech-language pathologist, 15 were mental-aged matched peers, and 15 were language-matched peers. The metalinguistic task required the participants to segment sentences and words to better understand awareness of words and word segmenting properties. For the sentence segmentation task, the researchers instructed the children to repeat "just a little bit" of the sentence. The

instruction utilized visual cues to demonstrate only repeating back part of the sentence, beads were removed from the string, along with the verbal reinforcement "you only said a little bit of the sentence." Following the training, the examiners asked the participants to repeat only part of sentences with the same prompt (i.e., "Only say a little bit."). For example, examiners gave participants the sentence, "We found a big tree," and prompted participants to divide the sentence at the word level by eliminating words from the sentence (i.e. "We found a big tree." becomes "We found a big.").

For the word segmentation task, examiners prompted participants to "only say a little bit of a word." Verbal instructions included "Now I am going to tell you something. It will be a word. I want you to say a little bit of it. If I say sailboat, then you say boat. If I say boat, you say boa." The researchers indicated that the participants were provided two examples and the puppet and bead demonstration was not needed. The word segmentation task included bisyllabic and monosyllabic words (i.e., "doctor" becomes "dock") and monosyllable words (i.e., "dock" becomes "do").

Following the segmentation tasks, examiners asked participants to define different types of words. For example, examiners asked participants to define what a word is (i.e., "What is a word?"), provide examples of long and short words (i.e., "Say a long word. What makes a word long?"; "Say a short word. What makes a word short?"), and provide examples of easy and hard words (i.e., "What is a hard word?"). Children were assigned scores based on their ability to demonstrate that words have meaning and are components within sentences, and that word complexity is determined by syllables and phonemes.

Therefore, if a child responded that, “Wood is a hard word, because wood is hard.” they received a lower score than if a child responded “Supercalifragilistic, because of the length and complexity of the word.”

Kamhi et al.’s results indicated that the children with language impairment had more difficulty segmenting the target items, and some could not make accurate segmentations at all. The mental-age matched peers demonstrated the strongest segmenting abilities; however, the language matched peers also performed significantly better than the children with language impairment. The results also demonstrated that children with language impairment performed significantly lower than both sets of peers when asked to define words. Generally, the children with language impairment had difficulty providing examples, as well as definitions. Kamhi et. al concluded that children with language impairment perform significantly lower on understanding of segments of language components and in their understanding of words by definition. These findings provide further support of the strong correlation between metalinguistic skills and language abilities. This study did not address grammar and semantic knowledge, or knowledge of grammatical morphemes that contribute to metalinguistic knowledge.

Fujiki et. al (1987) examined the impact of language impairment on grammatical awareness. The aim of the study was to determine if children who have language impairment make gains to eventually have the same grammatical skills as their age-matched peers. The study included 100 6-to 10-year-old children who were classified as having a language impairment or typically developing. Participants were identified as

having a language impairment using the Test of Language Development-Intermediate (Newcomer & Hammill, 1988) and the Clinical Evaluation of Language Functions Screening Test (Semel & Wiig 1980). All of the participants with language impairment also had previously qualified for special education services and had a nonverbal IQ score that was considered within normal limits. To examine grammatical awareness, examiners presented sentences with incorrect sentence structures and instructed the participants to correct the sentence word order (i.e., verb-subject-object, subject-object-verb, and verb-object-subject). For example, participants attempted to correct the sentence "Open he the door." Examiners coded participants' responses for accuracy and type of errors. Scores were analyzed across age groups and language abilities.

Study results revealed a significant increase in accuracy beginning at 8 years of age. Even in the group of children with language impairment, the older children performed significantly better on the task. However, the group of children with typical language skills consistently outperformed the children with language impairment. The results indicate that children with language impairment have increased difficulty with grammatical awareness tasks; however, the authors note that there were errors with the open-ended response types, as not all participant responses followed the instructions. The study also only addresses one type of skill; whereas, the researchers acknowledge that it takes significantly more cognitive skills to manipulate many different aspects of language. This suggests that the information obtained in this study may not be exhaustive

in understanding the relationship between language impairments and metalinguistic awareness.

Redmond & Rice (2001) examined the grammatical judgement abilities of 5- to 8-year-old children with specific language impairment (SLI) relative to language-matched and age-matched peers. The study included 57, 5- to 8-year-olds. Examiners assessed language skills using parent/teacher report, as well as the Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1981) scores, MLU scores, articulation screening scores, and the Test of Language Development (Newcomer & Hammill, 1988). To assess metalinguistic skills, participants judged if sentences were grammatically correct. Examiners presented finite and infinite sentences, in which some had grammatical errors and some did not. Examples of finite sentences used are, "The space guy fall into the pool" and "The space guy fell off a block." In comparison, nonfinite sentences used followed a "make+verb" format, such as "He made the space guy fell into the pool" and "He made the space guy fall off the block."

Following the judgement task, participants attempted to produce regular and irregular verb forms by describing what a creature was doing or what the examiner made a creature do. For example, examiners showed participants a picture of a creature throwing a ball and then a picture following the action. The examiner prompted the participant to produce the target, "He threw the ball." Examiners then prompted participants to generate sentences based on the examiners' manipulations of the creature

performing the action. An example of a target sentence is, "You made the creature throw the ball."

On the grammatical judgement task, Redmond and Rice (2001) found a significant difference in judgement and production of grammatical morphemes. The participants with SLI made significantly more errors identifying correct and incorrect grammatical forms. They were also more likely to produce infinitive forms when asked to produce verbs. These study results indicate that language abilities contribute to knowledge and correct identification of grammatical morphemes; however, there is not an expansive study of metalinguistic skills for the participants.

Results across studies indicate that children with language impairment have significant difficulty with grammatical judgements and production; however, it is still unclear at this time if this delay is consistent across all metalinguistic skills that contribute to semantic and morphological knowledge. Additionally, previous studies have not provided a complete analysis of age as a contributor to metalinguistic abilities. Previously, smaller age ranges (i.e., 3- to 6-years old, 6- to 10-years old) have been targeted to determine grammatical skills and word awareness; however, results indicated that metalinguistic skills significantly increase at 8 years of age (Fujiki et. al,1987). This study expands previous age ranges across semantic, syntactic, and morphological tasks to better understand the relationship between language abilities and metalinguistic skills.

Current Study

The current study aimed to determine if children who present with low language abilities demonstrate lower metalinguistic skills across a variety of metalinguistic tasks. This study differs from previous research as it targets a more extensive battery of metalinguistic awareness tasks. In addition to grammatical judgements, it also examines awareness of words and their meaning, and ability to manipulate grammatical morphemes. This study also differs from others in that children were identified based on language abilities, and not specific language diagnoses.

This study aimed to answer the following question:

- Does performance on a range of metalinguistic tasks differed based on participants' language ability and/or age and/or the specific task

Method

Participants

At the 2014 - 2018 Minnesota State Fairs, 1,096 participants completed the study protocol. The University of Minnesota, Twin Cities sponsors a *Driven to Discover* building dedicated to research in which the study team administered the 30-min protocol. The study team included the primary research investigator and several undergraduate and graduate research assistants. Participants included fairgoers who consented to participate in the study. All participants were between the ages of 3 and 10 years of age and were fluent in English. Birthdate and language information was obtained in a demographic survey. The primary investigator informed interested caregivers about the purpose of the study, what they and their children would do as part of the study, and that they could

cease participation at any time. The primary investigator obtained parent consent prior to the start of the assessments. The children received a small stuffed animal as an incentive for participating.

Participants at Risk for Language Impairment. Of the 1,096 participants, we identified 48 children as having low language (LL) abilities. We included these participants in the experimental group. Identification was based on the Clinical Evaluation of Language Fundamentals-Fourth Edition/Preschool (CELF-4/CELF-P; Semel, Wiig, & Secord, 2003) Recalling Sentences standard scores. The Recalling Sentences subtest is norm-referenced with a mean of 10 and a SD of 3. We included children who earned a standard score of 6 or lower (-1 SD) in the LL group. We excluded participants who spoke a second language and/or did not complete the Recalling Sentences subtest. We grouped the 48 participants who met this criterion by age (3- & 4-year Olds, $n = 15$; 5- & 6-year Olds, $n = 17$; 7- & 8-year Olds, $n = 16$). For further information of participant characteristics, refer to Table 1. Of these children, 10 parents reported their child having received speech-language pathology services or reading and writing services. Of these 10, 2 children were reported as having a diagnosis for Attention-Deficit/Hyperactivity Disorder and/or Autism Spectrum Disorder. An additional four children were reported as having Attention-Deficit/Hyperactivity Disorder, but were not receiving services.

Participants who are Typically Developing. From our sample of 1,096 participants, we created a control group of children who demonstrated typical

development. We excluded participants who scored lower than 1.0 standard deviations on the Recalling Sentences subtest of the CELF-4 or CELF-P (SS < 7) and/or the Kaufman Brief Intelligence Test-Second Edition-Matrices (KBIT-2; Kaufman & Kaufman, 2004) (SS < 85). We also excluded participants if parent questionnaire responses indicated that that their child had a history of and/or was receiving intervention for hearing loss, language impairments, behavioral diagnoses, reading and/or writing impairments, or seizures. Using the remaining sample, we matched each participant identified as having low language to a participant identified as developing typically based on chronological age gender, ethnicity, and household. Variance in age between matches that was accepted was +/- 8 months, while variance in annual household income that was accepted was +/- $75,000. If a participant could not be matched on a specific minority ethnicity, they were matched with a peer who identified as a different minority. Further participant characteristic information can be found in Table 1.

Table 1. Participant Group Characteristics

	3- & 4-Year Olds		5- & 6-Year Olds		7- & 8-Year Olds	
Characteristic	TD (*n* = 15)	LL (*n* = 15)	TD (*n* = 17)	LL (*n* = 17)	TD (*n* = 16)	LL (*n* = 16)
Age (months)						
Mean	49.44	50.44	74.18	74.71	94.56	94.06
SD	6.22	7.13	6.2	6.29	6.98	7.31
Min-Max	38-59	37-59	61-83	68-83	85-106	84-104
CELF[1] (standard score)						
Mean	11.33	4.47	11.65	5.29	11.69	4.75
SD	2.47	1.81	2.32	1.36	2.92	1.57
Min-Max	7-16	0-6	7-16	1-6	7-18	1-6
KBIT-2[2] (scaled score)						
Mean	13.40*	9.80*	108.35	92.24	110.30	94.63
SD	4.34*	3.49*	7.39	13.17	9.62	14.01
Min-Max	5-21*	2-15*	96-120	64-121	88-129	73-116
Gender						
Female:Male	10:6	10:6	5:12	5:12	9:7	9:7
Race						
White:Other	15:1	12:4	14:3	11:6	14:2	14:2
Caregiver Education						
High School	3	9	4	6	3	5
College	2	2	8	4	5	9
Graduate	7	3	4	6	8	4
Household Income (USD)						
$0-$50,000	4	3	3	5	2	4
$50,001-$150,000	9	10	12	10	10	8
$150,001+	3	3	2	2	4	4

Note. [1] Clinical Evaluation of Language Fundamentals-Fourth Edition/Preschool (CELF-4/CELF-P; Semel, Wiig, & Secord, 2003)
[2]Kaufman Brief Intelligence Test-Second Edition (KBIT-2; Kaufman & Kaufman, 2004)
*Raw Score Values

Assessments

Trained undergraduate and graduate research assistants administered and scored all assessments following standardized procedures. Assistants completed training on assessment administration prior to data collection and training on scoring protocols following data collection. Assessments included the Martices subtest of the Kaufman Brief Intelligence Test-Second Edition (KBIT-2; Kaufman & Kaufman, 2004), the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals-Fourth Edition/Preschool (CELF-4/CELF-P; Semel, Wiig, & Secord, 2003), and a four-task Metalinguistic Probe. Caregivers completed additional assessments, including a demographic survey.

KBIT-2 Matrices Subtest (Kaufman & Kaufman, 2004). The matrices subtest of the KBIT-2 measures nonverbal cognitive abilities. This subtest has a total of 42 items that assess problem solving, relationships between images, and ability to complete analogies. For example, for one item, the examiner presents the child a picture of a rabbit and asks which picture from a field of four demonstrates a relationship. For this item, the child should pick a carrot over a ball, bone, and shoe. The basal for the subtest is 3. The subtest was discontinued when a participant responded incorrectly to four consecutive items. The subtest is norm-referenced and yields a standard score with a mean of 100 and SD of 15.

CELF (Semel, Wiig, & Secord, 2003) Recalling Sentences Subtest. The Recalling Sentences subtest assesses expressive language skills. The participants'

performance on this assessment provided information on morphology, syntax, semantic skills, as well as long-term memory. The examiner presented the sentences one time. The participants were prompted to repeat the sentence exactly as it was heard. For example, they may be asked to repeat "The dad bought a book for his son who likes funny stories". The CELF-4 (Semel, Wiig, & Secord, 2003) comprises 32 sentences and the CELF-P (Semel, Wiig, & Secord, 2003) comprises 13 sentences. Participants received a score for each item. Scores were determined by the deviations from the original sentence. A sentence with no errors received a score three; whereas, a sentence with one error received a score of two, a sentence with two or three errors received a score of one, and a sentence with four or more errors was scored as a zero. The CELF-P (Semel, Wiig, & Secord, 2003) had two items that had a maximum possible score of two. The sentences on both versions increased in complexity. The subtest was discontinued when a participant received five consecutive scores of zero on the CELF-4 (Semel, Wiig, & Secord, 2003) and three consecutive scores of zero on the CELF-P (Semel, Wiig, & Secord, 2003). The subtest is norm-referenced and yields a standard score with a mean of 10 and SD of 3.

Metalinguistic Probe. The Metalinguistic Protocol consisted of four subtests, including a Word Manipulation Task, a Word Swap Task, a Morpheme Production Task, and a Grammatical Judgment Task. The research assistants administered the probe tasks via an iPad and recorded scores on a paper form. Individual tasks were discontinued when the child responded incorrectly on 5 consecutive items. Percent correct scores for

each metalinguistic task were obtained for comparison between TD and LL groups. Percent scores were calculated by dividing the total number of items correct by the total administered. 6- to 10-year old participants in 2017 and 2018 did not complete the Word Manipulation Task. Preliminary analysis revealed that children in this age group consistently performed at ceiling this task. Likewise, 3- to 4-year old participants in 2017 and 2018 did not complete the Grammatical Judgment Task. Analysis revealed that the demands of this task were too complex for typically developing children. See Table 2 for sample sizes for each task. Each task was based on probes included in previous studies of children's metalinguistic abilities. (Chaney, 1992; Piaget, 1929; Berko-Gleason, 1958; Bialystok, 1986).

Table 2. Sample Size for Completed Metalinguistic Tasks

		3- & 4-Year Olds		**5- & 6-Year Olds**		**7- & 8-Year Olds**	
Metalinguistic Task		TD Group (*n* = 15)	LL Group (*n* = 15)	TD Group (*n* = 17)	LL Group (*n* = 17)	TD Group (*n* = 16)	LL Group (*n* = 16)
Word Manipulation							
	n	15	15	14	15	11	12
Word Swap							
	n	15	15	17	17	16	16
Morpheme Production							
	n	14	14	17	17	16	16
Grammatical Judgement							
	n	11	14	15	16	16	16

The Word Manipulation Task was derived from Chaney's (1992) variation of Smith and Tager-Flusberg's (1982) word/referent differentiation task. This task assesses children's ability to assign new names to four known objects and understanding that the features of the object do not change with a new label. During this task, the examiner told the child that they were making up a new language and provided novel names while showing the child a picture of a common item. For example, the examiner told the child that they were changing the name of a "cow" to a "mib." The examiner then asked the child four questions to ensure that the child understood that they were just changing the name of the item and that the item's properties did not change by asking yes and no questions, such as "Do mibs moo?" and "Can you throw a mib?"

The Word Swap Task was modeled after the sun/moon problem that Piaget (1929) created. This task also assesses semantic knowledge and the ability to change labels. More so, the Word Swap test item assesses the child's ability to control impulses. The items within this subtest require manipulation of concepts and labels, and the child may answer incorrectly across the eight items if they are unable to inhibit impulsive responses. During this task, the examiner provided instruction that the words "sun" and "moon" were going to be switched, but the meaning was not going to change. The examiner explained that, "Suppose everyone in the world agreed that from now on we will call the sun the moon and the moon will be called the sun." Then the examiner asked the child, "What would that thing up in the sky be called when you go to bed?" The child should

respond, “The sun.” The examiner then asked, “What would the sky look like when you see this?” The correct response would be “dark.”

The Morpheme Production Task was based on Berko’s Wug Task (Berko-Gleason, 1958). This task assesses children’s knowledge of and ability to assign appropriate known grammatical morphemes to novel base words. This task contained 17 items that assessed 10 unique grammatical forms (e.g., plural –s, possessive –s, past tense –ed). During this task, the examiner showed the child a picture and prompted the child to produce the novel word plus appropriate English grammatical morpheme. For example, the examiner showed the child a picture of a blue bird-like creature and said, “This is a wug”. Then, the examiner showed the child another picture with two bird-like creatures and said, “Now there is another one. There are two ____”. The examiner then prompted the children to complete the sentence using the target word “wugs.”

The Grammatical Judgement Task was based on Bialystok’s (1986) judgement task. This task assesses children’s grammatical knowledge and their ability to process sentences to determine their accuracy. This task also assesses children’s ability to inhibit responding based on meaning, but rather only evaluate the grammatical structure of the sentences. The task contained 13 pre-recorded items; however, the participants who participated in 2017 and 2018 received 12 of the items. All participants listened to pre-recorded stimuli with Sennheiser headphones. The task included 3 sentences that were grammatically incorrect, but semantically sound (e.g., Yesterday, I climb a tree.), 3 sentences that were grammatically correct, and semantically correct (e.g., A girl who

sings is a singer.), 3 sentences that were that were grammatically and semantically incorrect (e.g., The shoe are teaching.), and 4 (2014- 2016) or 3 (2017-2018) sentences that were grammatically sound, but semantically incorrect (e.g., The car is reading.). For this task, the child should respond that the sentence "The car is reading" is grammatically correct, but "the girl are fast" is not.

Demographic Survey. Caregivers completed a 29-question demographic survey either on paper or via an iPad. The iPad survey was administered using Qualtrics (Qualtrics, 2018). Information obtained included, the child's age and grade. Additional information obtained included the languages that the children spoke and the percentage of time each language was spoken, as well as what language was spoken in the home. Information was also obtained regarding caregiver education level, household income, number of individuals in the household, and if the child has a diagnosis (e.g., Attention-Deficit/Hyperactivity Disorder or Autism Spectrum Disorder) and/or receives services related to language, cognition, or literacy skills.

Reliability

Experienced researcher assistants provided training to the student research assistants on the administration and scoring of the assessments to ensure reliability. Administration of the metalinguistic tasks and CELF-P/4 (Semel, Wiig, & Secord, 2003) were audio recorded to aid scoring. Research assistants scored all responses online, but could go back and listen to the recordings to ensure appropriate scoring. Each participant's audio file was identified by a unique, but de-identified participant number.

Participant numbers corresponded to all assessments and entries into a database. Trained research assistants entered scores into a database. Research assistants verified scoring as data was entered into the database. Additionally, different research assistants entered the assessments a second time. Entries were compared for accuracy and discrepancies were corrected.

Statistical Analyses

We conducted a 3 x 2 ANOVA to evaluate the effects of age and language status on metalinguistic performance. The age factor comprised three groups: 3- & 4-year Olds, 5- & 6-year Olds, and 7- & 8-year Olds. The language status factor comprised two groups: children with LL and those with TD. The primary dependent variables were the percentages of items correct for each metalinguistic task. We calculated η^2 for each main effect. Values below 0.13 have a small effect, values between 0.13 and 0.26 have a moderate effect, and values larger than 0.26 have a high effect (Cohen,1998). A higher effect size indicates that the relationship carries greater clinical significance. We completed follow-up analyses for each significant effect using the Tukey HSD procedure to control for Type I error across the pairwise comparisons. We also compared the LL and TD groups within each age group using *t*-tests and calculated effect sizes for the pairwise comparisons using Cohen's *d* (Cohen et. al, 1988). Cohen's *d* provides information on effect size of differences between means. A small effect size indicates that there is little clinical value; whereas, a large effect size is considered to be clinically

meaningful. Low values are considered to be around 0.2, medium values are 0.50, and large values are considered to be larger than 0.80 (Cohen, 1992).

Results

The means and standard deviations for metalinguistic performance across age and language groups are presented in Table 3. The ANOVA indicated a significant main effect for age, $F(8, 132) = 8.43$, $p < .001$, $\eta^2 = .34$, and a significant main effect for language status, $F(4, 66) = 6.93$, $p < .001$, $\eta^2 = .30$. Both of these effects were supported by large effect sizes. The interaction between age and language was not significant, $F(8, 132) = 1.85$, $p = .07$, $\eta^2 = .10$ with a small effect size. The post-hoc analysis for age revealed that for the Word Manipulation Task, the 7- & 8-year Olds outperformed the other two groups (ps < .001). For the Word Swap Task, the 7- & 8-year Olds outperformed the 3- & 4-year Olds ($p = .04$). For the Morpheme Production, the 7- & 8-year Olds outperformed the 6- & 7-year Olds ($p < .02$) and both groups outperformed the 3- & 4-year Olds (ps < .001). For the Grammatical Judgment Task, the 7- & 8-year Olds outperformed the other two groups (ps < .01). The TD Group outperformed the LL Group on all tasks (ps < .01) with exception of the Grammatical Judgment Task ($p = .20$).

We conducted *t*-tests and calculated effect sizes within age groups to compare the metalinguistic task scores of the LL and TD groups. The associated *p*- and *t*-values appear in Table 3. These comparisons are also displayed in Figure 1. Results indicated significant LL and TD group differences for the 3- & 4-years Olds on the Word Swap Task ($p = 0.02$); whereas, differences on the Word Manipulation Task ($p = 0.19$),

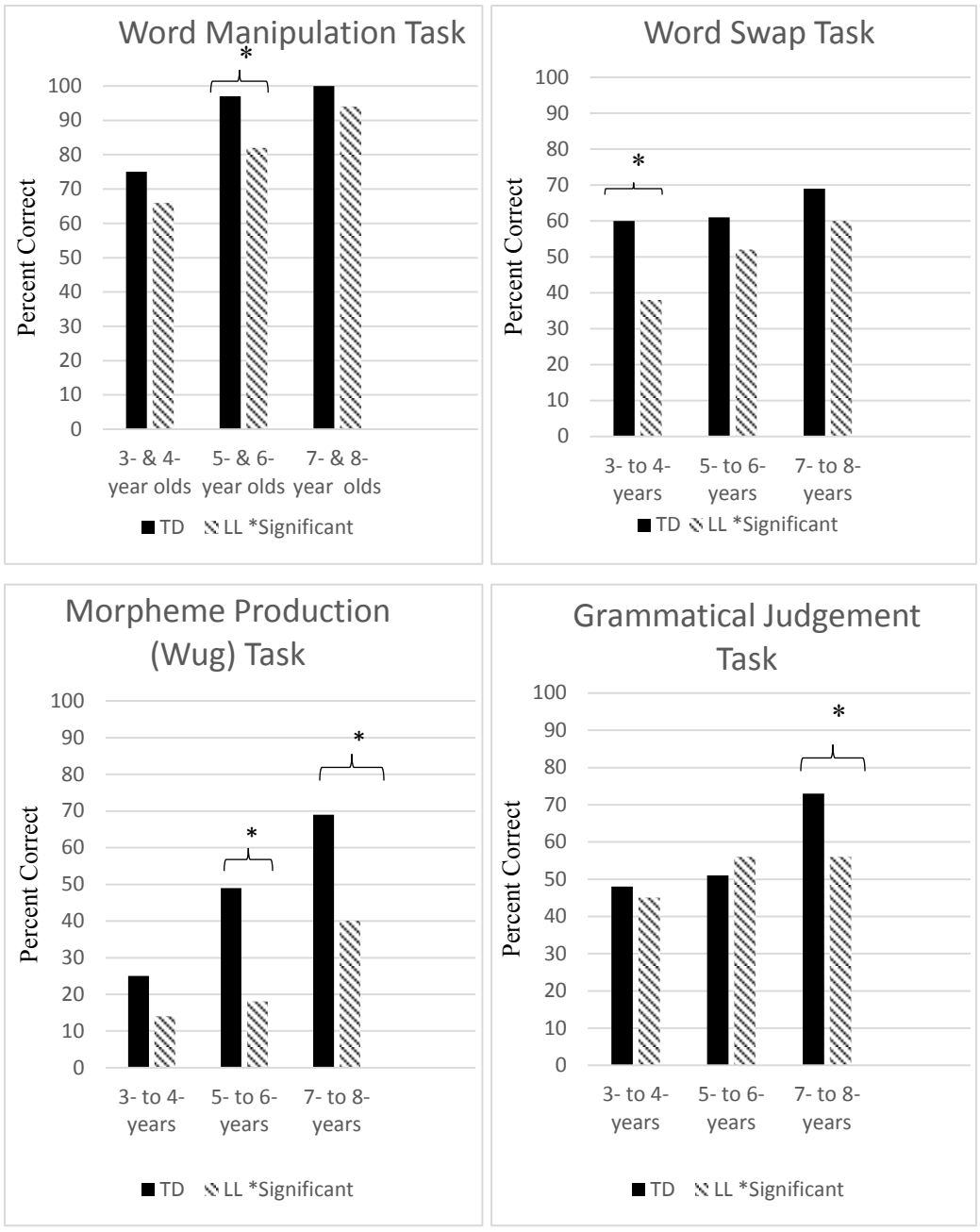
Morpheme Production Task ($p = 0.09$), and Grammatical Judgement Task ($p = 0.47$) were not significant. The 5- & 6-year Olds in the LL group demonstrated lower scores on the Word Manipulation Task ($p = 0.02$) and the Morpheme Production Task ($p < 0.01$) than the TD group; however, group differences were not significant for the Word Swap Task ($p = 0.14$) and the Grammatical Judgement Task ($p = 0.25$). The 7- & 8-year Olds in the LL group demonstrated lower scores on the Morpheme Production Task ($p < 0.01$) and the Grammatical Judgement Task ($p = 0.01$); however, there was not a significant difference between the TD and LL groups on the Word Manipulation Task ($p = 0.07$) or the Word Swap Task ($p = 0.17$). The findings for 3- & 4-year old children were supported by small effect sizes on the Word Manipulation and Grammatical Judgement Tasks, medium effect size on the Morpheme Production Task, and large effect size on the Word Swap Task. The 5- & 6-year Olds also demonstrated small effect sizes on the Word Manipulation and Grammatical Judgement Tasks, a medium effect size for the Word Swap Task, and a large effect size for the Morpheme Production Task. Further the 7- & 8-year Olds demonstrated a medium effect size for the Word Swap Task and large effect sizes for the Word Manipulation, Morpheme Production, and Grammatical Judgement Task.

Table 3. Participant Percent Correct by Age and Metalinguistic Task

	3- & 4-Year Olds		**5- & 6-Year Olds**		**7- & 8-Year Olds**	
Metalinguistic Task	TD Group (*n* = 15)	LL Group (*n* = 15)	TD Group (*n* = 17)	LL Group (*n* = 17)	TD Group (*n* = 16)	LL Group (*n* = 16)
Word Manipulation						
Mean	75.00	66.25	97.32	81.67	100.00	93.75
SD	20.46	21.50	72.37	22.96	< 0.01	10.99
p	0.26		0.02		0.07	
d	0.42		0.29		0.80	
Word Swap						
Mean	60.00	38.33	61.76	52.21	69.53	60.16
SD	22.76	19.17	17.38	19.38	20.90	16.59
p	0.01		0.14		0.17	
d	1.03		0.52		0.50	
Morpheme Production						
Mean	25.04	13.82	48.67	18.13	69.75	40.07
SD	20.82	20.07	17.10	18.84	18.91	27.02
p	0.09		< 0.01		< 0.01	
d	0.55		1.70		1.23	
Grammatical Judgement						
Mean	48.19	44.69	51.00	56.28	73.00	56.21
SD	12.46	17.54	13.07	11.65	20.66	14.35
p	0.47		0.25		0.01	
d	0.23		0.43		0.94	

Note. *P- and d-values based on t-test analyses*.

Figure 1. Percent of items correct by task, age, and language group.



Discussion

The current study aimed to examine the metalinguistic skills of 3- to 8-year old children with low language abilities across a battery of metalinguistic tasks. Compared to previous research, the current study expanded the metalinguistic tasks that were assessed, including a word swap task, word manipulation task, morpheme production task, and a grammatical judgement task. Results of the ANOVA suggest that children with low language ability have weaker metalinguistic skills compared to children developing typically and that older children have stronger metalinguistic skills than younger children. Pairwise comparisons revealed that effects varied across tasks. For the 3- & 4-year Olds, the only comparison that was significantly different was based on the Word Swap Task; although, the children with low language received lower scores across all tasks. The 5- & 6-year Olds with low language demonstrated significant differences relative to their TD peers on the Word Manipulation Task and the Morpheme Production Task. The children with low language also demonstrated lower scores that were not significant on the Word Swap Task. The 5- & 6-year Olds who presented with low language were the only group to outperform the children in the TD group on the Grammatical Judgement Task. The 7- & 8-year Olds that presented with low language also followed the trend to perform lower across metalinguistic tasks compared to children in the TD group. A significant difference was present for the Grammatical Judgment Task; however, differences on the Word Manipulation Task, Word Swap Task, and Morpheme Production Task were not significant.

The current study is consistent with previous research that reveals children with low language demonstrate low metalinguistic skills when compared to TD peers. Kahmi et. al (1985) found that 3- to 6-year old children that have a language impairment will perform lower than mental-matched peers on tasks that assess word awareness. Similarly, the current study identified that children in a similar age group with low language will also perform lower than typically developing children on tasks that assess word awareness (i.e., Word Manipulation Task and Word Swap Task). However, the current study expanded on previous tasks by also assessing processing and control.

Similarly, Kahmi et. al (1985) identified that typically developing children perform higher on tasks that assess their ability to segment words into different components. Redmond & Rice (2001) assessed the ability to make judgments and produce grammatical morphemes and found that children with low language make significantly more errors. While the current study assessed the participant's ability to add morphemes to alter meaning in the Morpheme Production Task, our results also suggest that children with low language demonstrate lower awareness of morphological structure.

On the Grammatical Judgement Task for which the 7- to 8-year old children with low language performed significantly lower than the typically developing peers. Results were consistent with Fujiki et. al (1987) who also found significant differences between 7- and 8-year old children with language impairment and TD children. Fujiki (1987) also identified significant differences for 6-year old participants; whereas, the current study

did not. Redmond and Rice (2001) also found that children with low language demonstrated difficulty identifying grammatical errors.

The current study also aimed to determine if differences in metalinguistic skills varied by task. As Bialystok outlined, metalinguistic skills consist of knowledge and ability to process linguistic features which will develop as language skills are obtained. Results from the current study suggest that differences in metalinguistic skills may be contingent upon the specific skill being assessed. The Word Manipulation Task and the Word Swap Task assess children's ability to process semantic features; whereas, the Morpheme Production Task and the Grammatical Judgement Task assess children's ability to process morphemes and syntax, respectively. Because of the difference in what the tasks are measuring, it was hypothesized that performance on tasks may vary. This was supported by the study's results. Children who presented with low language performed significantly less on some tasks; however, not on all of the tasks. The 3- & 4-year Olds performed higher on tasks assessing semantic features compared to syntactic features. On the grammatical judgement task, the 3- & 4-year Olds performed at chance level across both groups, suggesting the task demands were greater than syntactic skills typically acquired by 3- and 4-years old. The 5- & 6-year Olds performed similarly, participants performed higher on semantic tasks compared to syntactic tasks. Additionally, the 5- & 6-year Olds performed just over chance level across both groups. The 7- & 8-year Olds demonstrated a similar trend; however, the Word Manipulation Task was the only task to yield higher scores compared to the other three tasks. Further,

while the 3- & 4-year Olds and 5- & 6-year Olds performed at chance level on the Grammatical Judgement Task across both groups, the 7- to 8-year old TD children performed above chance, but the children that presented with low language performed at the same level as the 5- & 6-year Olds. These results support Bialystok's (1986, 2001) findings that metalinguistic skills are variable, and change over the course of development. It also highlights the importance of studying a battery of metalinguistic tasks, compared to a single skill as previously done.

We also aimed to determine if performance on the metalinguistic tasks was impacted by child age. Analyses revealed that performance across metalinguistic tasks improved with age, indicating that age is an important factor when analyzing metalinguistic skills. It was also observed that significant differences between children who presented with low language and children with TD varied between age groups. Specifically, while the 3- & 4-year Olds with low language had significantly lower scores than the TD group on the Word Swap Task, both 5- & 6- year Olds and 7- & 8-year Olds did not demonstrate significant differences. On the Morpheme Production Task, the 5- & 6- year Olds and 7- & 8-year Olds in the TD group significantly outperformed their peers in the LL group; however, the 3- & 4- year Olds performed relatively low across both groups (< 30% accuracy). This is consistent with Bialystok's (1989) findings that metalinguistic abilities are influenced by language development. The Morpheme Production Task may target skills that are too difficult for young children, even for children who typically developing. Similar results were found on the Grammatical

Judgment Task. The 3- & 4-year Olds and the 5- to 6-year Olds performed relatively low across both the TD and LL groups. This may indicate that the skills needed for the task are typically developed beyond 6 years of age. These results are consistent with previous studies. Fujiki et. al (1987) found significant differences on grammatical judgement tasks between typically developing children and children with language impairments in 6-, 7-, & 8-year olds, but not in the younger participants. Additionally, Smith (1982) identified a similar trend that 3- & 4-year old children will perform higher on semantic tasks compared to syntactic tasks.

Limitations

The current study identified children who present with low language on an expressive language task. This provided inclusionary criteria that was less restrictive than previous studies; and limits generalization of the results. Further assessment of receptive and expressive language may be warranted to identify specific areas of language that are weak and may impact metalinguistic skills. In future studies, greater details regarding the language profiles of children will aid in interpretation and generalization. This study was further limited by the small sample size. The current study included a wider range of participant ages; however, due to exclusionary factors, there was no data for children over 8-years of age. Future research should focus on expanding the sample size and further expand the age range included. This will continue to provide information on the link between metalinguistic skills and language development.

Further investigation is also warranted to develop methods for utilizing metalinguistic skills to improve language abilities. The current study adds to our understanding of the relationship between language and metalinguistic skills indicating that language profile impacts performance. Thus, the development and evaluation of interventions that target metalinguistic skills could improve semantic, syntactic, and morphologic language deficits in children. Intervention should focus on targeting both knowledge and processing of linguistic features to promote language skills.

Conclusion

This study further supports the link between metalinguistic skills and language. Findings suggest that 3- to 8-year old children who present with low language abilities will also have decreased metalinguistic skills when compared to their typically developing peers. It is important to consider how children process and use knowledge of linguistic features, as it aids in academic success and overall language outcomes (Kahmi et al.,1985). The link between language and metalinguistic skills provides further support for future research on how metalinguistic tasks can be utilized to both understand typical language development and target language skills in children with low language abilities. Future research should continue to provide information on the characteristics of this relationship, as well as to provide further evidence to support the use of metalinguistic skills in intervention.

Illustrations

Table 1. Participant Group Characteristics

	3- & 4-Year Olds		**5- & 6-Year Olds**		**7- & 8-Year Olds**	
Characteristic	TD (*n* = 15)	LL (*n* = 15)	TD (*n* = 17)	LL (*n* = 17)	TD (*n* = 16)	LL (*n* = 16)
Age (months)						
Mean	49.44	50.44	74.18	74.71	94.56	94.06
SD	6.22	7.13	6.2	6.29	6.98	7.31
Min-Max	38-59	37-59	61-83	68-83	85-106	84-104
CELF[1] (standard score)						
Mean	11.33	4.47	11.65	5.29	11.69	4.75
SD	2.47	1.81	2.32	1.36	2.92	1.57
Min-Max	7-16	0-6	7-16	1-6	7-18	1-6
KBIT-2[2] (scaled score)						
Mean	13.40*	9.80*	108.35	92.24	110.30	94.63
SD	4.34*	3.49*	7.39	13.17	9.62	14.01
Min-Max	5-21*	2-15*	96-120	64-121	88-129	73-116
Gender						
Female:Male	10:6	10:6	5:12	5:12	9:7	9:7
Race						
White:Other	15:1	12:4	14:3	11:6	14:2	14:2
Caregiver Education						
High School	3	9	4	6	3	5
College	2	2	8	4	5	9
Graduate	7	3	4	6	8	4
Household Income (USD)						
$0-$50,000	4	3	3	5	2	4
$50,001-$150,000	9	10	12	10	10	8
$150,001+	3	3	2	2	4	4

Note. [1] Clinical Evaluation of Language Fundamentals-Fourth Edition/Preschool (CELF-4/CELF-P; Semel, Wiig, & Secord, 2003)
[2]Kaufman Brief Intelligence Test-Second Edition (KBIT-2; Kaufman & Kaufman, 2004)
[*]Raw Score Values

Table 2. Sample Size for Completed Metalinguistic Tasks

	3- & 4-Year Olds		**5- & 6-Year Olds**		**7- & 8-Year Olds**	
Metalinguistic Task	TD Group (*n* = 15)	LL Group (*n* = 15)	TD Group (*n* = 17)	LL Group (*n* = 17)	TD Group (*n* = 16)	LL Group (*n* = 16)
Word Manipulation						
n	15	15	14	15	11	12
Word Swap						
n	15	15	17	17	16	16
Morpheme Production						
n	14	14	17	17	16	16
Grammatical Judgement						
n	11	14	15	16	16	16

Table 3. Participant Percent Correct by Age and Metalinguistic Task

	3- & 4-Year Olds		**5- & 6-Year Olds**		**7- & 8-Year Olds**	
Metalinguistic Task	TD Group (*n* = 15)	LL Group (*n* = 15)	TD Group (*n* = 17)	LL Group (*n* = 17)	TD Group (*n* = 16)	LL Group (*n* = 16)
Word Manipulation						
Mean	75.00	66.25	97.32	81.67	100.00	93.75
SD	20.46	21.50	72.37	22.96	< 0.01	10.99
p	0.26		0.02		0.07	
d	0.42		0.29		0.80	
Word Swap						
Mean	60.00	38.33	61.76	52.21	69.53	60.16
SD	22.76	19.17	17.38	19.38	20.90	16.59
p	0.01		0.14		0.17	
d	1.03		0.52		0.50	
Morpheme Production						
Mean	25.04	13.82	48.67	18.13	69.75	40.07
SD	20.82	20.07	17.10	18.84	18.91	27.02
p	0.09		< 0.01		< 0.01	
d	0.55		1.70		1.23	
Grammatical Judgement						
Mean	48.19	44.69	51.00	56.28	73.00	56.21
SD	12.46	17.54	13.07	11.65	20.66	14.35
p	0.47		0.25		0.01	
d	0.23		0.43		0.94	

Note. *P- and d-values based on t-test analyses*.

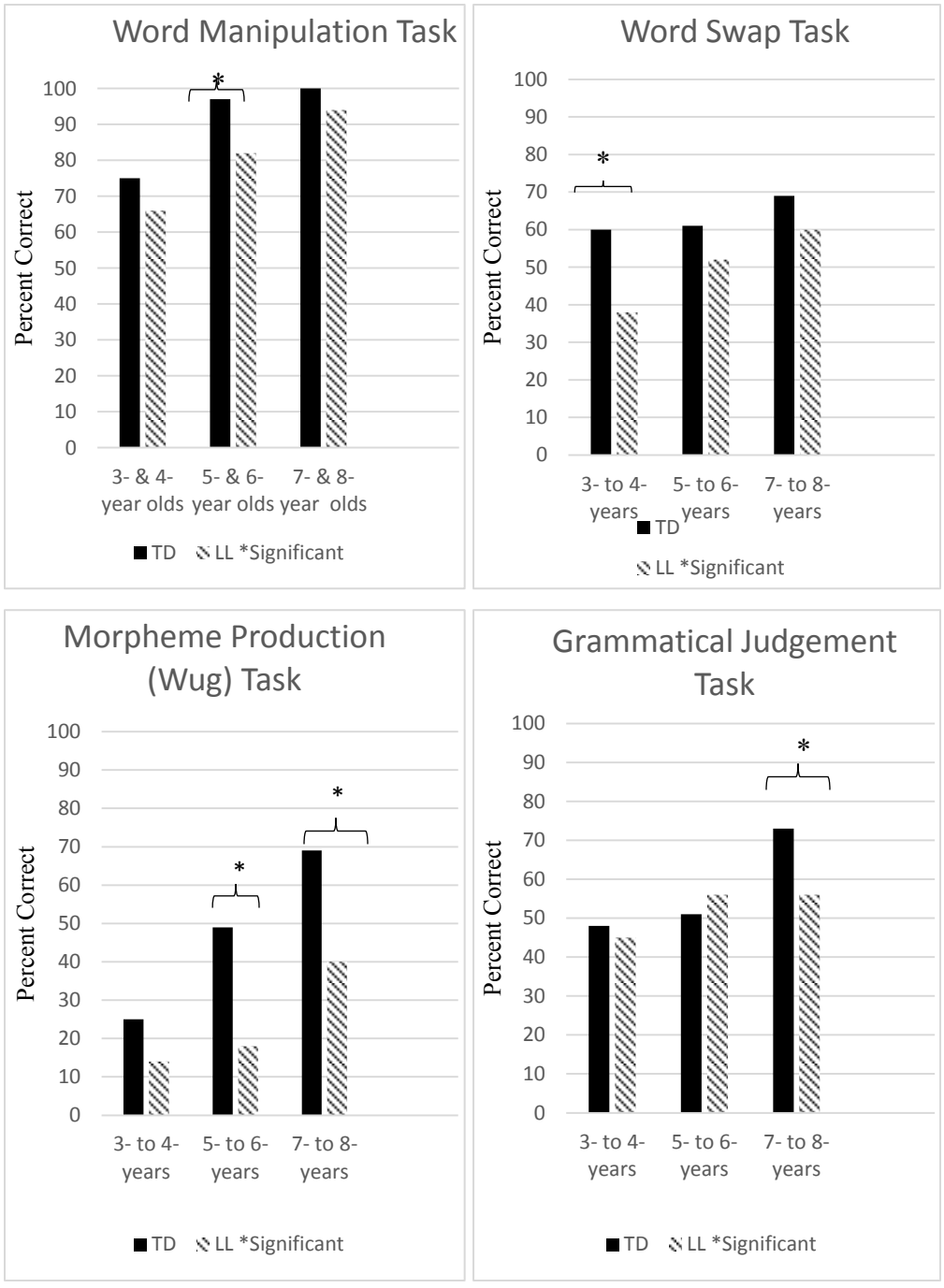


Figure 1. Percent of item correct by task, age, and language group.

Bibliography

Berko-Gleason, J. (1958). Children's learning of English morphology'. *Word*, *14*(2-3), 150-177.

Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annual review of applied linguistics*, *21*, 169-181.Bialystok, E., & Ellen, B. (Eds.). (1991). *Language processing in bilingual children*. Cambridge University Press.

Bialystok, E. (1987). Influences of bilingualism on metalinguistic development. *Interlanguage studies bulletin (Utrecht)*, *3*(2), 154-166.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child development*, 498-510.

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Catalise-2 Consortium, Adams, C., ... & Boyle, C. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068-1080.Cairns, H. S., & Mcdaniel, D. (n.d.). *Development of a Metalinguistic Skill: Judging the Grammaticality of Sentences*. *Communication Disorders Quarterly* (Vol. 27).

Chaney C. Language development, metalinguistic skills, and print awareness in 3-year-old children. Applied Psycholinguistics. 1992;13:485–514.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Statistical power analysis for the behavioral sciences: Cohen, J. (1988). (2nd ed.).

Hillsdale,NJ: Lawrence Erlbaum Associates, Publishers. 567 pages.

(1990). *Computers, Environment and Urban Systems*, *14*(1), 71.

Dunn, A., & Dunn, A. (1981). Peabody Picture Vocabulary Test–Revised. Circle Pines, MN: American Guidance Service.

Fujiki, M., Brinton, B., & Dunton, S. (1987). *THE ABILITY OF NORMAL AND LANGUAGE-IMPAIRED CHILDREN TO PRODUCE GRAMMATICAL CORRECTIONS*. *J*. *COMMUN*. *DISORD* (Vol. 20).

Kamhi, A. G. (1987). Metalinguistic abilities in language—impaired. *Topics in Language Disorders*.

Kamhi, A. G., Lee, R. F., & Nelson, L. K. (1985). *WORD, SYLLABLE, AND SOUND AWARENESS IN LANGUAGE-DISORDERED CHILDREN*. *Journal of Speech and Hearing Disorders* (Vol. 50).

Kaufman, A. S. (2004). Kaufman Brief Intelligence Test–Second Edition (KBIT-2). *Circle Pines, MN: American Guidance Service*.

Newcomer, P. L., & Hammill, D. D. (1988). Test of Language Development–Primary (2nd ed.). Austin, TX: Pro-Ed.

Papandropoulou, I., & Sinclair, H. (1974). What is a word?. *Human development*, *17*(4), 241-258.

Piaget, J. (1929). The child's concept of the world. *Londres, Routldge & Kegan Paul*.

Redmond, S. M., & Rice, M. L. (2001). *15Redmond*. *Journal of Speech, Language, and Hearing Research* • (Vol. 44). Retrieved from

Semel, E., Wiig, E. H., & Secord, W. A. (2003). Clinical Evaluation of Language Fundamentals

(CELF-4). the psychological corporation. *San Antonio, TX*.

Semel, E., and Wiig, E. (1980). Clinical Evaluation of Language Functions. Co- lumbus. Ohio: Charles E. Merrill.

Smith, C. L., & Tager-Flusberg, H. (1982). Metalinguistic awareness and language development. *Journal of Experimental Child Psychology*, *34*(3), 449-468.

Wiig, E. H., Secord, W., & Semel, E. M. (2004). *CELF preschool 2: clinical evaluation of language fundamentals preschool*. Pearson/PsychCorp.